

Method of hand language translation through a intermediate mode language

Publication number: CN1464433

Publication date: 2003-12-31

Inventor: GAO WEN (CN); MA JIYONG (CN); WANG CHUNLI (CN)

Applicant: CHINA ACADEMY OF SCIENCES INST (CN)

Classification:

- International: G06F17/28; G06F17/28; (IPC1-7): G06F17/28

- European:

Application number: CN20021021369 20020617

Priority number(s): CN20021021369 20020617

Also published as:

 CN1246793C (C)

[Report a data error here](#)

Abstract of CN1464433

The invention discloses a method of hand language translation through a intermediate mode language comprising, gathering hand language words data, collecting characteristic information in the hand language words data, identifying continuous sentences in the hand language according the characteristic information, recording the identification result of the intermediate mode language data, transforming the intermediate mode language data into the non-hand language words for output, and collecting non-hand language words data, based on the corresponding relationship between the intermediate mode language data and the non-hand language, converting the non-hand language words into the intermediate mode language data and recording. According to the intermediate mode language data, finding the corresponding hand language word data in the hand language word library, and combining the hand language words data into hand language image information for output.

.....
Data supplied from the esp@cenet database - Worldwide

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.⁷
G06F 17/28



[12] 发明专利申请公开说明书

[21] 申请号 02121369.0

[43] 公开日 2003 年 12 月 31 日

[11] 公开号 CN1464433A

[22] 申请日 2002.6.17 [21] 申请号 02121369.0

[71] 申请人 中国科学院计算技术研究所

地址 100080 北京市海淀区中关村南大街 8 号

[72] 发明人 高文 马继勇 王春立 吴江琴
陈熙林 宋益波 尹宝才 王兆其
山世光 曾炜 晏洁 吴枫
姚鸿勋 张洪明 吕岩 王瑞

[74] 专利代理机构 北京同立钧成知识产权代理有限公司

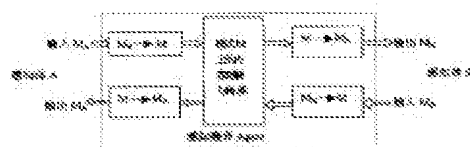
代理人 刘芳

权利要求书 5 页 说明书 20 页 附图 2 页

[54] 发明名称 通过中间模式语言进行手语翻译的方法

[57] 摘要

一种通过中间模式语言进行手语翻译的方法，包括：采集手语词语数据，提取该手语词语数据中的特征信息，根据该特征信息进行手语连续语句识别，然后记录中间模式语言数据的识别结果将中间模式语言数据转换为该非手语语言词语并输出；以及，采集非手语语言词语数据，根据中间模式语言数据与该非手语语言的对应关系，将该非手语语言词语转换为中间模式语言数据并记录，根据该中间模式语言数据，在手语词语库中找到相应的手语词语数据，再将该手语词语数据合成为手语图像信息输出；本发明将手语以及非手语模式的语言均与中间模式语言相对应，有利于手语翻译系统的扩展，方便了非手语语言与手语之间的相互转换。



1、一种通过中间模式语言进行手语翻译的方法，其特征在于：它包括：
将手语通过一中间模式语言数据形式翻译为非手语语言的方法；以及将非手语
5 语言通过一中间模式语言数据形式翻译为手语的方法。

2、根据权利要求 1 所述的通过中间模式语言进行手语翻译的方法，其特征
在于：将手语通过一中间模式语言数据形式翻译为非手语语言的具体步骤包
括：

步骤 101：采集手语词语数据；

10 步骤 102：提取该手语词语数据中的特征信息；

步骤 103：根据该特征信息进行手语连续语句识别，然后记录中间模式语
言数据的识别结果；

步骤 104：根据中间模式语言数据与相应的非手语语言的对应关系，将中
间模式语言数据转换为该非手语语言词语并输出。

15 3、根据权利要求 1 所述的通过中间模式语言进行手语翻译的方法，其特征
在于：将非手语语言通过一中间模式语言数据形式翻译为手语的具体步骤包
括：

步骤 201：采集非手语语言词语数据；

步骤 202：根据中间模式语言数据与该非手语语言的对应关系，将该非手
20 语语言词语转换为中间模式语言数据并记录；

步骤 203：根据该中间模式语言数据，在手语词语库中找到相应的手语词
语数据，再将该手语词语数据合成为手语图像信息输出。

4、根据权利要求 2 或 3 所述的通过中间模式语言进行手语翻译的方法，
其特征在于：手语翻译为非手语语言或非手语语言翻译为手语语言的方法还包
25 括：在采集手语词语数据或非手语数据的同时，还采集相应的人脸信息；然后，
提取该人脸信息中的特征数据，最后利用该特征数据在翻译输出时合成输出
的人脸图像。

5 5、根据权利要求 2 所述的通过中间模式语言进行手语翻译的方法，其特征在于：采集手语词语数据的具体方法为：采用数据手套采集手的各关节的传感数据；采用位置跟踪器输入手语手势的位置和方向数据；其中，数据手套装设在人体的左右手上；位置跟踪器包括一个发射器和一个以上的接收器；该发射器发出电磁波，接收器装设在人体的左右腕部，该接收器接收该电磁波并计算该接收器相对于发射器的位置和方向数据。

6、根据权利要求 2 所述的通过中间模式语言进行手语翻译的方法，其特征在于：提取该手语词语数据中的特征信息的具体方法为：

10 计算左右两手相对于参照的位置和方向，对手的各关节传感数据的每个分量进行归一化处理，并将处理后的数据作为 HMM 的训练样本，建立手语样本模型库。

7、根据权利要求 2 所述的通过中间模式语言进行手语翻译的方法，其特征在于：手语连续语句识别的具体方法为：用 Viterbi 译码方法计算测试样本与各种可能的模型序列的似然概率，将概率值最大的模型序列所对应的词序列
15 作为识别结果。

8、根据权利要求 7 所述的通过中间模式语言进行手语翻译的方法，其特征在于：手语连续语句识别的具体方法还包括基于统计的语言模型，在 Viterbi 搜索过程中嵌入二阶马尔科夫链，即语句 $W = w_1, w_2, \dots, w_n$ 的先验概率 $P(W)$ 可用下式计算：

20
$$P(W) = P(w_1, w_2, \dots, w_n) = \prod_{i=1}^n P(w_i | w_{i-1});$$

其中，

W 为被识别的语句；

w_1, w_2, \dots, w_n 为被识别的语句 W 中的各字；

$P(w_i | w_{i-1})$ 为字对的出现频度。

25 9、根据权利要求 3 所述的通过中间模式语言进行手语翻译的方法，其特

征在于：手语词语数据合成为手语图像信息的具体方法为：

采用 VRML 的人体表示模型建立虚拟人；

确定该虚拟人各自由度的角度值；

计算虚拟人每个肢体的位置和方向，确定出虚拟人的一个姿态；

3 忽略该姿态的非上肢关节角度；

按照规定的时间间隔连续显示一个手语运动中的每一个姿态，生成相应的手语运动图像。

10、根据权利要求 9 所述的通过中间模式语言进行手语翻译的方法，其特征在于：在生成手语运动图像时，还进一步在手语运动图像的相邻帧之间进行

10 平滑插值；具体的插值根据如下的公式计算：

$$G_i(t_r) = G_i(t_i) + \frac{t_r - t_i}{t_f - t_i} (G_i(t_f) - G_i(t_i))$$

其中，

t_1 和 t_2 分别为一个手语运动中两个相邻图像帧；

t_i 和 t_f 分别为 t_1 和 t_2 距起始点的时间值；

15 t_r 为被插入帧距起始点的时间值；

t_i 为起始点的时间值；

$G_i(t_r)$ 为被插入的自由度曲线函数值；

$G_i(t_i)$ 为起始点的自由度曲线函数值；

$G_i(t_f)$ 为 t_1 的自由度曲线函数；

20 $G_i(t_f)$ 为 t_2 的自由度曲线函数值。

11、根据权利要求 9 所述的通过中间模式语言进行手语翻译的方法，其特征在于：在生成手语运动图像时，还进一步采用基于四元组的运动插值方法对不连续手语帧中的复杂关节进行平滑过渡处理，具体的平滑处理根据如下的公式计算：

25
$$q_{i,r} = \frac{\sin(1-t_r)\theta}{\sin\theta} q_{i,i} + \frac{\sin(t_r\theta)}{\sin\theta} q_{i,f}$$

其中,

f_1 和 f_2 分别为一个手语运动中两个相邻手势的图像帧;

t_1 和 t_2 分别为 f_1 和 f_2 距起始点的时间值;

q_1 和 q_2 分别为关节在 t_1 , t_2 时刻的方向;

5 t_r 为插入帧距起始点的时间值;

θ 为.....并且由 $q_1 \cdot q_2 = \cos \theta$ 确定。

12、根据权利要求 4 所述的通过中间模式语言进行手语翻译的方法, 其特征在于: 所述的提取该人脸信息中的特征数据至少包括: 正面人脸特征的检测和侧面人脸特征点的检测; 其中,

10 正面人脸特征的检测至少包括: 面部特征粗定位、关键特征点检测和基于变形模板的特征形状检测;

侧面人脸特征点检测至少包括: 人脸侧面轮廓线的提取和人脸侧面特征点的检测。

13、根据权利要求 12 所述的通过中间模式语言进行手语翻译的方法, 其特征
15 在于: 所述的面部特征粗定位为: 首先定位虹膜的位置, 然后根据虹膜中心点的位置数据、面部器官结构的统计先验数据和面部灰度分布特性获得人脸其他器官的位置数据。

14、根据权利要求 12 所述的通过中间模式语言进行手语翻译的方法, 其特征
20 在于: 人脸关键特征点的检测为: 获取眼角点、嘴角点、下巴曲线上的主要特征点, 作为相应的器官模板参数的初值; 具体包括: 眼睛关键点的检测, 嘴部关键点的检测和下巴关键点的检测; 其中: 眼睛关键点包括左右眼角点和上下眼皮的界限点; 嘴部关键点包括两个嘴角点、上唇最高点和下唇最低点; 下巴的关键点包括左右嘴角的延长线与下巴的交点、过中唇点的垂线与下巴的交点、过左右嘴角点的垂线与下巴的交点、过左嘴角点往左下 45 度直线与下巴
25 的交点、过右嘴角点往右下 45 度直线与下巴的交点。

15、根据权利要求 12 所述的通过中间模式语言进行手语翻译的方法, 其

特征在于：基于变形模板的特征形状检测包括：对眼睛区域特征形状进行检测，获得眼睛模板参数；对嘴部形状进行检测，获得嘴部模板参数；对下巴形状进行检测，获得下巴模板参数。

16、根据权利要求 12 所述的通过中间模式语言进行手语翻译的方法，其特征
5 特征在于：所述的人脸侧面轮廓线的提取为：利用人脸的肤色特征将人脸区域分割出来；然后采用边缘检测，并根据人脸轮廓的先验数据定位轮廓线。

17、根据权利要求 12 所述的通过中间模式语言进行手语翻译的方法，其特征
10 特征在于：所述的人脸侧面特征点的检测为：以鼻尖点为界，将人脸轮廓线分为上下两段；通过曲线拟合，得到轮廓线的近似函数表达式，计算该函数一阶导数为零的点，将该点作为人脸侧面特征点。

18、根据权利要求 4 所述的通过中间模式语言进行手语翻译的方法，其特征
15 特征在于：利用特征数据合成输出人脸图像的方法为：在人脸模型上定义一个以上个特征点，将其当作对一般人脸模型的变形参数；对一般人脸中性模型进行整体变换，完成面部整体轮廓的修改，使其与特定人脸形和五官的位置相匹配；
得到具有特定人脸特征的中性三维人脸网格体。

19、根据权利要求 4 或 18 所述的通过中间模式语言进行手语翻译的方法，其特征
20 特征在于：利用特征数据合成输出人脸图像的方法还包括：根据唇部特征数据建立参数化的唇动模型，再根据唇动模型和语言发音的相应口型最后合成相应口型；具体为：

唇部模型采用两条抛物线拟合上唇线，一条抛物线拟合下唇线；选择两个嘴角点、两上唇抛物线的最高点、下唇抛物线的最低点、两上唇抛物线的交点，并在下唇抛物线上增加两点，上唇抛物线上增加两点，在两个嘴角点的连线上增加三组每两个可重合的点；张嘴时，各对重合点分离，分别成为上下内唇抛物线上的各点；并且，唇部的外轮廓的抛物线方程满足：

25
$$y = a(x - b)^2 + c$$

其中，a, b, c 可以通过将上述唇线各点的坐标值带入上述方程求解获得。

通过中间模式语言进行手语翻译的方法

技术领域:

- 5 本发明涉及一种通过中间模式语言进行手语翻译的方法，特别是指一种将手语通过一中间模式语言数据形式翻译为非手语语言以及将非手语语言通过一中间模式语言数据形式翻译为手语的方法。

背景技术:

- 语言是人们相互之间进行交流不可或缺的工具，但是，当今世界的正在使用的语言种类数以百计，如果加上地方方言和聋人的手语等，其数量更是难以统计；如此众多的语言种类使得使用不同语言的人群之间的交流变得非常困难，不仅在生理健康的人群之间是这样，在一些生理上具有残疾的人群中，这一问题尤其突出；因此，各种语言之间的翻译一直是困扰着全人类的一大问题。

- 随着人类科学技术的不断进步，尤其是在近 20 年内，计算机技术的迅猛发展，使得利用计算机将一种感知语言翻译成其他语言已成为了现实。但是，目前利用计算机进行语言翻译通常是仅仅将一种模式的语言直接翻译成另一种模式的语言；这种翻译的方法及其系统具有如下的缺点：

- 在计算机自动翻译领域中，一种被翻译语言与一种翻译语言之间通常只有一种固定的词汇对应关系。因此，现有的翻译系统无法实现对一种被翻译语言做多种语言的翻译；如果以上述的翻译系统实现从一种被翻译语言到多种语言的翻译，就需要在被翻译语言与每一种翻译语言之间均建立相应固定的词汇对应关系；这样，一方面会使得设计和实现相应的翻译系统的工作量巨大，另一方面翻译系统的语言扩展也不易实现。

发明内容:

- 25 本发明的主要目的在于提供一种通过中间模式语言进行手语翻译的方法，将手语翻译成一中间模式语言，再将该中间模式的语言进一步翻译成所需的非手语语言形式；或者将非手语语言转换为中间模式语言，然后再将该中间模式

语言翻译为手语。

本发明的另一目的在于提供一种通过中间模式语言进行手语翻译的方法，手语以及非手语模式的语言均与中间模式语言相对应，有利于手语翻译系统的扩展，以实现非手语语言与手语之间的相互转换。

5 本发明的目的是这样实现的：

一种通过中间模式语言进行手语翻译的方法，包括：将手语通过一中间模式语言数据形式翻译为非手语语言的方法；以及将非手语语言通过一中间模式语言数据形式翻译为手语的方法。

将手语通过一中间模式语言数据形式翻译为非手语语言的具体步骤包括：

10 步骤 101：采集手语词语数据；

步骤 102：提取该手语词语数据中的特征信息；

步骤 103：根据该特征信息进行手语连续语句识别，然后记录中间模式语言数据的识别结果；

15 步骤 104：根据中间模式语言数据与相应的非手语语言的对应关系，将中间模式语言数据转换为该非手语语言词语并输出。

将非手语语言通过一中间模式语言数据形式翻译为手语的具体步骤包括：

步骤 201：采集非手语语言词语数据；

步骤 202：根据中间模式语言数据与该非手语语言的对应关系，将该非手语语言词语转换为中间模式语言数据并记录；

20 步骤 203：根据该中间模式语言数据，在手语词语库中找到相应的手语词语数据，再将该手语词语数据合成为手语图像信息输出。

手语翻译为非手语语言或非手语语言翻译为手语语言的方法还包括：在采集手语词语数据或非手语数据的同时，还采集相应的人脸信息；然后，提取该人脸信息中的特征数据，最后利用该特征数据在翻译输出时合成输出的人脸图

25 像。

采集手语词语数据的具体方法为：采用数据手套采集手的各关节的传感数

据数据；采用位置跟踪器输入手语手势的位置和方向数据；其中，数据手套装设在人体的左右手上；位置跟踪器包括一个发射器和一个以上的接收器；该发射器发出电磁波，接收器装设在人体的左右腕部，该接收器接收该电磁波并计算该接收器相对于发射器的位置和方向数据。

- 5 提取该手语词语数据中的特征信息的具体方法为：计算左右两手相对于参照的位置和方向，对手的各关节传感数据的每个分量进行归一化处理，并将处理后的数据作为隐马尔可夫模型（Hidden Markov Model，简称 HMM）的训练样本，建立手语样本模型库。

如上所述的一个 HMM 可用参数： $\lambda=(A,B,\pi)$ 表示，

- 10 其中， $A=\{a_{ij}\}$ 为状态转移概率矩阵，

并且满足公式： $a_{ij}=P[q_{t+1}=S_j|q_t=S_i]$ ， $1\leq i,j\leq N$ ；

并满足约束条件： $a_{ij}\geq 0$ ， $1\leq i,j\leq N$ ， $\sum_{j=1}^N a_{ij}=1$ ， $1\leq i\leq N$ ；

上式中， N 为模型的状态数；

$\pi=\{\pi_i\}$ ， π_i 表示从第 i 个状态结点开始的概率，

- 15 并且满足公式： $\pi_i=P[q_1=S_i]$ ， $1\leq i\leq N$ ；

并满足约束条件： $\pi_i\geq 0$ ， $\sum_{i=1}^N \pi_i=1$ ， $1\leq i\leq N$ ；

$B=\{b_j(k)\}$ 为观测信号的概率密度，当观察符号为连续矢量时， $b_j(k)$ 是连续概率密度函数，并且： $b_j(k)=\sum_{m=1}^M c_{jm}G[\mu_{jm},\Sigma_{jm},O_k]$ ， $1\leq j\leq N$ ；

- 其中， N 为模型的状态数， M 为混合项数， O_k 为 k 时刻的观察向量； c_{jm} 为
20 混合比（Mixing proportion），并且满足：

$$\sum_{m=1}^M c_{jm}=1, 1\leq j\leq N, c_{jm}\geq 0, 1\leq j\leq N, 1\leq m\leq M$$

其中： G 取为高斯概率密度函数， μ_{jm} 和 Σ_{jm} 分别为高斯混合概率密度中第 m 个分量的均值向量和协方差矩阵；

$$G[\mu_{jm}, \Sigma_{jm}, O_k] = \frac{1}{\sqrt{2\pi} |\Sigma_{jm}|^2} \exp[-\frac{1}{2} (O_k - \mu_{jm})^T \Sigma_{jm}^{-1} (O_k - \mu_{jm})]$$

对应于同一个手势词的 K 组训练数据 $O = [O^{(1)}, O^{(2)}, \dots, O^{(K)}]$, 其中 $O^{(k)} = [O_1^{(k)}, O_2^{(k)}, \dots, O_{T_k}^{(k)}]$ 是第 k 组训练数据, T_k 是第 k 组训练数据的总帧数;

π_i, a_{ij} 和 c_{jm}, μ_m, Σ_m 的重估公式为:

$$\pi_i = \frac{1}{K} \sum_{k=1}^K \gamma_i^{(k)}(1)$$

即在时刻 $t=1$ 时, 状态结点 i 出现的期望概率;

$$\bar{a}_{ij} = \sum_{k=1}^K \sum_{t=1}^{T_k-1} \xi_i^{(k)}(t, j) / \sum_{k=1}^K \sum_{t=1}^{T_k-1} \gamma_i^{(k)}(t), \text{ 其中,}$$

分子部分为从状态结点 i 转移到状态结点 j 的期望概率;

分母部分为从状态结点 i 转移的期望概率;

10

$$\bar{c}_{jm} = \sum_{k=1}^K \sum_{t=1}^{T_k} \gamma_i^{(k)}(j, m) / \sum_{k=1}^K \sum_{t=1}^{T_k} \sum_{m=1}^M \gamma_i^{(k)}(j, m), \text{ 其中,}$$

分子部分为在状态结点 j 的第 m 个分支出现的期望概率;

分母部分为在状态结点 j 出现的期望概率;

15

$$\bar{\mu}_m = \sum_{k=1}^K \sum_{t=1}^{T_k} \gamma_i^{(k)}(j, m) \cdot O_t^{(k)} / \sum_{k=1}^K \sum_{t=1}^{T_k} \gamma_i^{(k)}(j, m), \text{ 其中,}$$

分子部分为在状态结点 j 第 m 个分支出现观察序列 O 的期望概率;

分母部分为在状态结点 j 第 m 个分支出现的期望概率;

$$\bar{\Sigma}_m = \sum_{k=1}^K \sum_{t=1}^{T_k} \gamma_i^{(k)}(j, m) \cdot (O_t^{(k)} - \mu_m)(O_t^{(k)} - \mu_m)^T / \sum_{k=1}^K \sum_{t=1}^{T_k} \gamma_i^{(k)}(j, m), \text{ 其中,}$$

20

分子部分为在状态结点 j 第 m 个分支出现观察序列 O 的均方差的期望概率;

分母部分为在状态结点 j 第 m 个分支出现的期望概率。

手语连续语句识别的具体方法为:

模型库建好以后, 用韦特比 (Viterbi) 译码方法计算测试样本与各种可能的

25 模型序列的似然概率, 概率值最大的模型序列所对应的词序列即为识别结果。

设词汇表的容量为 V ，词的模型编号为 $k=1, \dots, V$ ，对应的模型参数为 $(\pi^k, A^k, c^k, \mu^k, U^k)$ ，每个词的状态数均为 L ，输入手语帧序列编号为 $i=1, 2, \dots, N$ ；

当词汇在同一模型内转移时 ($j > 1$):

$$\begin{aligned} 5 \quad & \Pr(j, k) = p(i, j, k) * \max\{\Pr(j, k) * A_{j,j}^k, \Pr(j-1, k) * A_{j-1,j}^k\} \\ & T(i, j, k) = k \\ & F(i, j, k) = j_0 = \arg \max\{\Pr(j, k) * A_{j,j}^k, \Pr(j-1, k) * A_{j-1,j}^k\} \end{aligned}$$

当词汇在模型的边界($j=1$)发生转移时:

$$\begin{aligned} 10 \quad & \Pr(j, k) = p(i, j, k) * \max\{\Pr(j, k) * A_{j,j}^k, \Pr(L, k') * \Pr(k | k'), 1 \leq k' \leq V\} \\ & T(i, j, k) = k_0 = \arg \max\{\Pr(j, k) * A_{j,j}^k, \Pr(L, k') * \Pr(k | k'), 1 \leq k' \leq V\} \\ & F(i, j, k) = j_0 = \arg \max\{\Pr(j, k) * A_{j,j}^k, \Pr(L, k') * \Pr(k | k'), 1 \leq k' \leq V\} \end{aligned}$$

开始时 ($i=1$)， $\Pr(1, k) = p(1, 1, k)$ ， $\Pr(j, k) = 0, j > 1$

$$\begin{aligned} 15 \quad & T(1, j, k) = -1 \\ & F(1, j, k) = -1 \end{aligned}$$

可用上述公式递归地求出各个 $\Pr(L, k)$ ，从而可得到全局最大概率:

$$\Pr_{\max} = \max_k \{\Pr(L, k)\}$$

由 $T(i, j, k)$ 和 $F(i, j, k)$ 回溯出最佳路径 (T_i, F_i) (倒序):

$$\begin{aligned} 20 \quad & T_N = \arg \max_k \{\Pr(L, k)\} \\ & F_N = L \\ & T_i = T(i+1, T_{i+1}, F_{i+1}), \quad N-1 \geq i \geq 1 \\ & F_i = F(i+1, T_{i+1}, F_{i+1}), \quad N-1 \geq i \geq 1 \end{aligned}$$

得到识别结果; 其中,

25 $p(i, j, k)$: 为在词 k 的第 j 个状态出现第 i 帧的概率;

$\Pr(j, k)$: 为到当前输入帧 i 为止, 从开始经状态转移到词 k 的第 j 个状态的最大概率;

$T(i, j, k)$: 记录前一帧所在的模型的序号;

$F(i, j, k)$: 记录前一帧在模型 $T(i, j, k)$ 中所处的状态。

为了提高识别精度,在 Viterbi 搜索过程中嵌入二阶马尔科夫链 (Bigram),
即:语句的先验概率可用下式计算:

$$P(W) = P(w_1, w_2, \dots, w_n) = \prod_{i=1}^n P(w_i | w_{i-1})$$

其中,

5 W 为被识别的语句;

w_1, w_2, \dots, w_n 为被识别的语句 W 中的各字;

$P(w_i | w_{i-1})$ 为字对的出现频度。

手语词语数据合成为手语图像信息的具体方法为:

10 采用 VRML 的人体表示模型建立虚拟人;

确定该虚拟人各自由度的角度值;

计算虚拟人每个肢体的位置和方向,确定出虚拟人的一个姿态;

忽略该姿态的非上肢关节角度;

按照规定的时间间隔连续显示一个手语运动中的每一个姿态,生成相应的

15 手语运动图像。

在生成手语运动图像时,还进一步在手语运动图像的相邻帧之间进行平滑插值;具体的插值根据如下的公式计算:

$$G_i(t_f) = G_i(t_i) + \frac{t_f - t_i}{t_f - t_i} (G_i(t_f) - G_i(t_i))$$

其中,

20 t_i 和 t_f 分别为一个手语运动中两个相邻图像帧;

t_i 和 t_f 分别为 t_i 和 t_f 距起始点的时间值;

t_f 为被插入帧距起始点的时间值;

t_i 为起始点的时间值;

$G_i(t_f)$ 为被插入的自由度曲线函数值;

25 $G_i(t_i)$ 为起始点的自由度曲线函数值;

$G_i(t_{f_i})$ 为 f_i 的自由度曲线函数;

$G_i(t_{f_i})$ 为 f_i 的自由度曲线函数值。

在生成手语运动图像时, 还进一步采用基于四元组的运动插值方法对不连续手语帧中的复杂关节进行平滑过渡处理, 具体的平滑处理根据如下的公式计

5 算:

$$q_{t_r} = \frac{\sin(1-t_r)\theta}{\sin\theta} q_{f_1} + \frac{\sin(t_r\theta)}{\sin\theta} q_{f_2}$$

其中,

f_1 和 f_2 分别为一个手语运动中两个相邻手势的图像帧;

t_{f_1} 和 t_{f_2} 分别为 f_1 和 f_2 距起始点的时间值;

10 q_{f_1} 和 q_{f_2} 分别为关节在 t_{f_1} , t_{f_2} 时刻的方向;

t_r 为插入帧距起始点的时间值;

θ 为.....并且由 $q_{f_1} \cdot q_{f_2} = \cos\theta$ 确定。

所述的提取该人脸信息中的特征数据至少包括: 正面人脸特征的检测和侧面人脸特征点的检测; 其中,

15 正面人脸特征的检测至少包括: 面部特征粗定位、关键特征点检测和基于变形模板的特征形状检测;

侧面人脸特征点检测至少包括: 人脸侧面轮廓线的提取和人脸侧面特征点的检测。

20 所述的面部特征粗定位为: 首先定位虹膜的位置, 然后根据虹膜中心点的位置数据、面部器官结构的统计先验数据和面部灰度分布特性获得人脸其他器官的位置数据。

人脸关键特征点的检测为: 获取眼角点、嘴角点、下巴曲线上的主要特征点, 作为相应的器官模板参数的初值; 具体包括: 眼睛关键点的检测, 嘴部关键点的检测 and 下巴关键点的检测; 其中: 眼睛关键点包括左右眼角点和上下眼皮的界限点; 嘴部关键点包括两个嘴角点、上唇最高点和下唇最低点; 下巴的

25

关键点包括左右嘴角的延长线与下巴的交点、过中唇点的垂线与下巴的交点、过左右嘴角点的垂线与下巴的交点、过左嘴角点往左下 45 度直线与下巴的交点、过右嘴角点往右下 45 度直线与下巴的交点。

基于变形模板的特征形状检测包括：对眼睛区域特征形状进行检测，获得
5 眼睛模板参数；对嘴部形状进行检测，获得嘴部模板参数；对下巴形状进行检测，获得下巴模板参数。

所述的人脸侧面轮廓线的提取为：利用人脸的肤色特征将人脸区域分割出来；然后采用边缘检测，并根据人脸轮廓的先验数据定位轮廓线。

所述的人脸侧面特征点的检测为：以鼻尖点为界，将人脸轮廓线分为上下
10 两段；通过曲线拟合，得到轮廓线的近似函数表达式，计算该函数一阶导数为零的点，将该点作为人脸侧面特征点。

利用特征数据合成输出人脸图像的方法为：在人脸模型上定义多个特征点。该特征点可以从特定人的正面和侧面像中提取出来，特征点的自动提取属于人脸面部图象检测与分析范畴；假定已经应用分析与识别技术从特定人脸图
15 象上提取了所需特征或变形曲线，然后将其当作对一般人脸模型的变形参数。一般人脸中性模型是一个三维网格体，每个特征点的三维坐标是已知的，在从一般人脸中性模型到特定人脸中性模型的修改过程中，要进行两种变换；首先：要对一般人脸中性模型进行整体变换，完成面部整体轮廓的修改，使其与特定人的脸形和五官的大致位置相匹配。设变换前的人脸模型上点坐标为 (x, y, z) ，
20 变换后为 (x', y', z') ，变换前后脸的中心点坐标分别为 $o(x_0, y_0, z_0)$ 和 $o'(x'_0, y'_0, z'_0)$ ，其中脸的中心点 o 定义为两眼眼角点连线及人脸纵向中轴线之间的交点。参数 $p, q1, q2$ 分别定义为中心点到鬓角，中心点到额头中点，中心点到下巴之间的距离；参数 u 定义为嘴部中心点到耳朵下缘距离；参数 $r1, r2, r3$ 在侧面分别定义为额头最高处到发际的距离，中心点到耳朵上缘的距离，嘴角点到耳朵下边缘的距离。
25

对于上半部分脸（眼睛水平线以上的部分）采用如下修改公式：

$$\begin{aligned}x' &= x_0' + \frac{p'}{p}(x - x_0) \\y' &= y_0' + \frac{q'}{q_1}(y - y_0) \\z' &= z_0' + \frac{r'}{r_1}(z - z_0)\end{aligned}$$

对于脸的中间部分和下部可以类似地修改。

- 5 局部变换中的眼睛区域上点的修改公式如下：

设 (x, y, z) 为变换前眼睛区域点的坐标， (x', y', z') 为变换后眼睛区域点的坐标，则有：

$$\begin{aligned}x' &= ax + by + cz \\y' &= dx + ey + fz\end{aligned}$$

- 10 其中， a, b, c, d, e, f 可通过带入变换前后的三组特征点，解六元线性方程组得到；经过这个变换，可以达到对眼睛微小位置的移动及其形状的改变。从上式看出，修改仅发生在 x 和 y 方向上，而没有进行深度方面的修改，因此不能很好地反映特征人脸的侧面信息。同理可进行眉毛和嘴部的修改。

- 15 设变换前后鼻子区域点的坐标分别为 (x, y, z) 和 (x', y', z') ，鼻子的中心点坐标分别为 (x_0, y_0, z_0) 和 (x'_0, y'_0, z'_0) ，对鼻子部分的变换依照如下公式计算：

$$\begin{aligned}x' &= x_0' + \frac{p'}{p}(x - x_0) \\y' &= y_0' + \frac{q'}{q}(y - y_0) \\z' &= z_0' + \frac{r'}{r}(z - z_0)\end{aligned}$$

- 20 完成了整体变换和局部变换之后，就得到了一张基本具有特定人脸特征的中性三维人脸网格体。

- 利用特征数据合成输出人脸图像的方法还包括：唇部模型采用两条抛物线拟合上唇线，一条抛物线拟合下唇线。选择两个嘴角点，两上唇抛物线的最高点，下唇抛物线的最低点和两上唇抛物线的交点；另外，分别在下唇抛物线上、上唇抛物线上、在两个嘴角点的连线上增加若干个点，分别成为上下内唇抛物线上的各点。唇部的外轮廓的抛物线则满足如下公式：
- 25

$$y = a(x - b)^2 + c$$

其中，系数 a, b, c 可以通过将已知点的坐标带入上述方程即可解得。

动态的唇动模型利用上述张口模型，描述为五条相互关联起来的抛物线，其中包括上唇两条，下唇一条，上下内唇各一条。在进行参数驱动的唇动合成中，给出纵、横方向上开口的距离，或者给出嘴角点以及上下唇的最高点，即可确定相应的唇部开口状态。

- 5 本发明提供一种通过中间模式语言进行手语翻译的方法，将手语及非手语翻译成一中间模式语言，再将该中间模式的语言进一步翻译成所需的语言形式；手语以及非手语模式的语言均与中间模式语言相对应，有利于手语翻译系统的扩展，方便了非手语语言与手语之间的相互转换。

以下结合附图和具体的实施例对本发明做详细的进一步说明：

10 附图说明：

图 1 为本发明的基本原理示意图。

图 2 为本发明中将手语翻译为中间模式语言的流程示意图。

图 3 为本发明中将中间模式语言翻译为手语的流程示意图。

图 4 为本发明一实施例的整体流程结构示意图。

15 具体实施方式：

参见图 1，本发明的基本原理是：以中间模式语言 M 作为手语 M_h 和非手语语言 M_g 之间转换的必要途径；即：将手语 M_h 翻译为非手语语言 M_g 或将非手语语言 M_g 翻译为手语 M_h ，均通过一中间模式语言 M 进行。

实施例 1：手语转变为语音输出

- 20 参见图 2 和图 4，将手语通过一中间模式语言数据形式翻译为非手语语言的具体方法为：

- 首先，采集手语词语数据；本发明的一实施例中，采用两只具有 18 个传感器的数据手套及其配套设备-位置跟踪器作为手势输入设备，该位置跟踪器由一个发射器和若干个接收器构成，发射器发出电磁波，每个接收器接收该电磁波，然后根据接收到的电磁波计算该接收器相对于发射器的位置和方向数据。在作出手语动作的人体左右手腕上各配一个接收器，因为发射器的位置不
- 25

固定,测试时采集的手语数据的坐标经常会发生变化,所以还要把第三个接收器放在人身体上,用该接收器的位置作为参照点和参照坐标系,通过参照该第三接收器的位置和方向数据,获取左、右手上的接收器相对参照坐标系的位置和方向为不变量特征。

- 5 提取该手语词语数据中的特征信息的具体方法为:计算左右两手相对于参照的位置和方向,对手的各关节传感数据的每个分量进行归一化处理,并将处理后的数据作为HMM训练样本,建立手语样本模型库。

如上所述的一个HMM可用参数: $\lambda=(A,B,\pi)$ 表示,

其中, $A=\{a_{ij}\}$ 为状态转移概率矩阵,

- 10 并且满足公式: $a_{ij}=P[q_{i+1}=S_j|q_i=S_i], 1\leq i,j\leq N$;

并满足约束条件: $a_{ij}\geq 0, 1\leq i,j\leq N, \sum_{j=1}^N a_{ij}=1, 1\leq i\leq N$;

上式中, N 为模型的状态数;

$\pi=\{\pi_i\}$, π_i 表示从第 i 个状态结点开始的概率,

并且满足公式: $\pi_i=P[q_1=S_i], 1\leq i\leq N$;

- 15 并满足约束条件: $\pi_i\geq 0, \sum_{i=1}^N \pi_i=1, 1\leq i\leq N$;

$B=\{b_j(k)\}$ 为观测信号的概率密度,当观察符号为连续矢量时, $b_j(k)$ 是连续概率密度函数,并且: $b_j(k)=\sum_{m=1}^M c_{jm}G[\mu_{jm},\Sigma_{jm},O_k], 1\leq j\leq N$;

其中, N 为模型的状态数, M 为混合项数, O_k 为 k 时刻的观察向量; c_{jm} 为混合比(Mixing proportion),并且满足:

- 20
$$\sum_{m=1}^M c_{jm}=1, 1\leq j\leq N, c_{jm}\geq 0, 1\leq j\leq N, 1\leq m\leq M$$

其中: G 取为高斯概率密度函数, μ_{jm} 和 Σ_{jm} 分别为高斯混合概率密度中第 m 个分量的均值向量和协方差矩阵;

$$G[\mu_{jm},\Sigma_{jm},O_k]=\frac{1}{\sqrt{2\pi}|\Sigma_{jm}|^{\frac{1}{2}}}\exp\{-\frac{1}{2}(O_k-\mu_{jm})^T\Sigma_{jm}^{-1}(O_k-\mu_{jm})\}$$

对应于同一个手势词的 K 组训练数据 $O = [O^{(1)}, O^{(2)}, \dots, O^{(K)}]$, 其中 $O^{(k)} = [O_1^{(k)}, O_2^{(k)}, \dots, O_{T_k}^{(k)}]$ 是第 k 组训练数据, T_k 是第 k 组训练数据的总帧数;

π_i, a_{ij} 和 c_{jm}, μ_m, Σ_m 的重估公式为:

$$\pi_i = \frac{1}{K} \sum_{k=1}^K \gamma_i^{(k)}(1)$$

5 即在时刻 $t=1$ 时, 状态结点 i 出现的期望概率;

$$\bar{a}_{ij} = \sum_{k=1}^K \sum_{t=1}^{T_k-1} \xi_{ij}^{(k)}(t, t+1) / \sum_{k=1}^K \sum_{t=1}^{T_k-1} \gamma_i^{(k)}(t), \text{ 其中,}$$

分子部分为从状态结点 i 转移到状态结点 j 的期望概率;

分母部分为从状态结点 i 转移的期望概率;

$$10 \quad \bar{c}_{jm} = \sum_{k=1}^K \sum_{t=1}^{T_k} \gamma_j^{(k)}(t, m) / \sum_{k=1}^K \sum_{t=1}^{T_k} \sum_{m=1}^M \gamma_j^{(k)}(t, m), \text{ 其中,}$$

分子部分为在状态结点 j 的第 m 个分支出现的期望概率;

分母部分为在状态结点 j 出现的期望概率;

$$\bar{\mu}_m = \sum_{k=1}^K \sum_{t=1}^{T_k} \gamma_j^{(k)}(t, m) \cdot O_t^{(k)} / \sum_{k=1}^K \sum_{t=1}^{T_k} \gamma_j^{(k)}(t, m), \text{ 其中,}$$

15 分子部分为在状态结点 j 第 m 个分支出现观察序列 O 的期望概率;

分母部分为在状态结点 j 第 m 个分支出现的期望概率;

$$\bar{\Sigma}_m = \sum_{k=1}^K \sum_{t=1}^{T_k} \gamma_j^{(k)}(t, m) \cdot (O_t^{(k)} - \mu_m)(O_t^{(k)} - \mu_m)' / \sum_{k=1}^K \sum_{t=1}^{T_k} \gamma_j^{(k)}(t, m), \text{ 其中,}$$

分子部分为在状态结点 j 第 m 个分支出现观察序列 O 的均方差的期望概率;

20 分母部分为在状态结点 j 第 m 个分支出现的期望概率。

在提取手语词语数据的特征信息后, 再根据该特征信息进行手语连续语句识别, 然后记录中间模式语言数据的识别结果; 具体的方法为:

25 采用半连续隐马尔可夫模型 (Semi-Continuous Hidden Markov Model, 简称 SCHMM) 分别处理手形, 位置和方向, 以便减少码本的数目, 然后通过建立位置, 方向和手形的多维字母串来描述手语。

首先, 用单数据流对所有词建立连续的模型, 对所有状态结点上的均值向

量的左右手形、右手相对左手位置、右手相对左手方向和三个接收器之间的距离等六部分数据分别聚类，聚类的步骤如下：

初始化：从训练向量集合中任意选出多个向量作为初始码本；

寻找最接近的码字：对每个训练向量，在当前码本中寻找与之最接近的码字向量，并把这个向量分配到该码字所对应的集合中；

码字的修正：将码字修正为其所对应的集合中所有训练向量的均值；

上述的两步骤，直至各类中向量的均方差低于给定阈值。

得到聚类中心后，再对所有模型的每个状态结点的均值向量进行量化。每个状态结点只记录与之距离最近的码本的序号。在识别时，对每帧识别数据只需计算它与各码本的距离，用与该结点所纪录的码本的距离来代替与该结点的均值向量的距离。

上述的模型库建好以后，用韦特比 (Viterbi)译码方法计算测试样本与各种可能的模型序列的似然概率，概率值最大的模型序列所对应的词序列即为识别结果。

15 设词汇表的容量为 V ，词的模型编号为 $k=1, \dots, V$ ，对应的模型参数为 $(\pi^k, A^k, c^k, \mu^k, U^k)$ ，每个词的状态数均为 L ，输入手语帧序列编号为 $i=1, 2, \dots, N$ ；

当词汇在同一模型内转移时 ($j > 1$):

$$\begin{aligned} \Pr(j, k) &= p(i, j, k) * \max\{\Pr(j, k) * A_{j,j}^k, \Pr(j-1, k) * A_{j-1,j}^k\} \\ 20 \quad T(i, j, k) &= k \\ F(i, j, k) &= j_0 = \arg \max\{\Pr(j, k) * A_{j,j}^k, \Pr(j-1, k) * A_{j-1,j}^k\} \end{aligned}$$

当词汇在模型的边界 ($j=1$) 发生转移时:

$$\begin{aligned} \Pr(j, k) &= p(i, j, k) * \max\{\Pr(j, k) * A_{j,j}^k, \Pr(L, k^*) * \Pr(k | k^*), 1 \leq k^* \leq V\} \\ 25 \quad T(i, j, k) &= k_0 = \arg \max\{\Pr(j, k) * A_{j,j}^k, \Pr(L, k^*) * \Pr(k | k^*), 1 \leq k^* \leq V\} \\ F(i, j, k) &= j_0 = \arg \max\{\Pr(j, k) * A_{j,j}^k, \Pr(L, k^*) * \Pr(k | k^*), 1 \leq k^* \leq V\} \end{aligned}$$

开始时 ($i=1$)， $\Pr(1, k) = p(1, 1, k)$ ， $\Pr(j, k) = 0, j > 1$

$$T(1, j, k) = -1$$

$$F(1, j, k) = -1$$

可用上述公式递归地求出各个 $\text{Pr}(L, k)$ ，从而可得到全局最大概率：

$$\text{Pr}_{\max} = \max_k \{\text{Pr}(L, k)\}$$

5 由 $T(i, j, k)$ 和 $F(i, j, k)$ 回溯出最佳路径 (T_i, F_i) (倒序)：

$$T_N = \arg \max_k \{\text{Pr}(L, k)\}$$

$$F_N = L$$

$$T_i = T(i+1, T_{i+1}, F_{i+1}), \quad N-1 \geq i \geq 1$$

$$F_i = F(i+1, T_{i+1}, F_{i+1}), \quad N-1 \geq i \geq 1$$

10 得到识别结果；其中，

$p(i, j, k)$ ：为在词 k 的第 j 个状态出现第 i 帧的概率；

$\text{Pr}(j, k)$ ：为到当前输入帧 i 为止，从开始经状态转移到词 k 的第 j 个状态的最大概率；

$T(i, j, k)$ ：记录前一帧所在的模型的序号；

15 $F(i, j, k)$ ：记录前一帧在模型 $T(i, j, k)$ 中所处的状态。

为了提高识别精度，在 Viterbi 搜索过程中嵌入二阶马尔科夫链 (Bigram)，即：语句的先验概率可用下式计算：

$$P(W) = P(w_1, w_2, \dots, w_n) = \prod_{i=1}^n P(w_i | w_{i-1})$$

其中，

20 W 为被识别的语句；

w_1, w_2, \dots, w_n 为被识别的语句 W 中的各字；

$P(w_i | w_{i-1})$ 为字对的出现频度。

由此获得即可实现将手语的动作信息转换为中间模式语言的语句。

当需要将输入的手语翻译为非手语语句时 (例如：语音语句)，则根据中
25 间模式语言数据与相应的非手语语言的对应关系，将中间模式语言数据转换为
该非手语语言词语并输出。在本实施例，作为手语到非手语之间的中间模式

语言为文字文本；以语音输出为例：当需要语音输出时，可利用语音合成将采用文本中的文字表达出来。其中，语音合成可通过将孤立词的音频数据作简单的平滑处理后进行连接播出。具体的语音合成方法则采用现有的以文本文件合成为语音的系统。

5 实施例 2：语音转变为手语输出

参见图 3 和图 4，将非手语语言通过一中间模式语言数据形式翻译为手语的具体方法为：

首先，采集非手语语言词语数据，利用现有的语音识别技术将口语转换为中间模式语言数据；本实施例中，该语音识别可利用 IBM 公司开发的语音数据
10 开发工具实现。

然后，根据中间模式语言数据与该非手语语言的对应关系，将该非手语语言词语转换为中间模式语言数据并记录；本实施例中，所采用的中间模式语言存储为文字文本。

最后，根据该文字文本，在手语词语库中找到相应的手语词语数据，再将
15 该手语词语数据合成为手语图像信息输出；具体的实现包括：文本的输入、分析、切分、以文本表示的自然语言到手语码的转换以及手语词语数据合成为手语图像信息等步骤。

其中，文本的输入、分析、切分可利用现有的自然语言识别方法实现，然后将切份后的自然语言词语与手语词库中相应的手语词语相对应，获得该手语
20 词语的手语特征数据。该手语特征数据用于最后合成相应的手语图像。

上述手语图像合成的具体方法为：

首先，采用虚拟现实建模语言(Virtual Reality Modeling Language, 简称 VRML)的人体表示模型建立虚拟人；确定该虚拟人各自由度的角度值，并计算虚拟人每个肢体的位置和方向，确定出虚拟人的一个姿态；

25 由于手语是人体上肢运动，手语运动是人体运动在人体上肢关节上的投影，因此，显示手语（即将手语映射到虚拟人姿态）时，可将一个手语运动表

示扩充为一个完整的人体运动表示；也就是说，手语的显示可以使用人体运动显示的通用方法进行，因此，通过忽略上述步骤获得的虚拟人姿态的非上肢关节角度，就可以获得该虚拟人的一个手语姿态；或者说通过忽略上述步骤获得的虚拟人姿态的非上肢关节角度，就可以将虚拟人的一个手语姿态以一个完整

5 的人体运动姿态表示。

在获得所有的手语运动姿态数据以后，再按照规定的时间间隔连续显示一个手语运动中的每一个姿态，生成相应的手语运动图像。

在生成手语运动图像时，还进一步在手语运动图像的相邻帧之间进行平滑插值；具体的插值根据如下的公式计算：

$$10 \quad G_i(t_f) = G_i(t_1) + \frac{t_f - t_1}{t_2 - t_1} (G_i(t_2) - G_i(t_1))$$

其中，

t_1 和 t_2 分别为一个手语运动中两个相邻图像帧；

t_{f1} 和 t_{f2} 分别为 t_1 和 t_2 距起始点的时间值；

t_f 为被插入帧距起始点的时间值；

15 t_i 为起始点的时间值；

$G_i(t_f)$ 为被插入的自由度曲线函数值；

$G_i(t_1)$ 为起始点的自由度曲线函数值；

$G_i(t_{f1})$ 为 t_1 的自由度曲线函数；

$G_i(t_{f2})$ 为 t_2 的自由度曲线函数值。

20 在生成手语运动图像时，还进一步采用基于四元组的运动插值方法对不连续手语帧中的复杂关节进行平滑过渡处理，具体的平滑处理根据如下的公式计算：

$$q_{t_f} = \frac{\sin(1-t_f)\theta}{\sin\theta} q_{f1} + \frac{\sin(t_f\theta)}{\sin\theta} q_{f2}$$

其中，

25 t_1 和 t_2 分别为一个手语运动中两个相邻手势的图像帧；

t_1 和 t_2 分别为 f_1 和 f_2 距起始点的时间值;

q_1 和 q_2 分别为关节在 t_1 , t_2 时刻的方向;

t_r 为插入帧距起始点的时间值;

θ 为.....并且由 $q_1 \cdot q_2 = \cos \theta$ 确定.

5

以上的两个实施例中仅给出了手语运动的识别和合成方法。事实上,手语语言中通常还包含着人体面部表情信息和唇动信息;并且,在对手语进行识别或者合成时,往往还需要对具体的表达方的面部表情特征和唇动信息给予描述。

10 参见图 4,在本发明的实施例中还进一步对表达方特定的人脸进行检测获取相应的人脸特征及唇动信息,并且在输出端进行合成,与手语语句同步输出。具体的人脸及唇动信息的检测和合成的方法如下:

首先,提取该人脸信息中的特征数据,它至少包括:正面人脸特征的检测和侧面人脸特征点的检测;其中,

15 正面人脸特征的检测至少包括:面部特征粗定位、关键特征点检测和基于变形模板的特征形状检测;

侧面人脸特征点检测至少包括:人脸侧面轮廓线的提取和人脸侧面特征点的检测。

20 所述的面部特征粗定位为:首先定位人眼虹膜的位置,然后根据两个虹膜中心点的位置数据、面部器官结构的统计先验数据和面部灰度分布特性获得该人脸其他器官的位置数据。

人脸关键特征点的检测为:获取眼角点、嘴角点、下巴曲线上的主要特征点,作为相应的器官模板参数的初值;具体包括:眼睛关键点的检测,嘴部关键点的检测 and 下巴关键点的检测;其中:眼睛关键点包括左右眼角点和上下眼皮的界限点;嘴部关键点包括两个嘴角点、上唇最高点和下唇最低点;下巴的关键点包括左右嘴角的延长线与下巴的交点、过中唇点的垂线与下巴的交点、

25

过左右嘴角点的垂线与下巴的交点、过左嘴角点往左下 45 度直线与下巴的交点、过右嘴角点往右下 45 度直线与下巴的交点。

基于变形模板的特征形状检测包括：对眼睛区域特征形状进行检测，获得眼睛模板参数；对嘴部形状进行检测，获得嘴部模板参数；对下巴形状进行检测，获得下巴模板参数。

所述的人脸侧面轮廓线的提取为：利用人脸的肤色特征将人脸区域分割出来；然后采用边缘检测，并根据人脸轮廓的先验数据定位轮廓线。

所述的人脸侧面特征点的检测为：以鼻尖点为界，将人脸轮廓线分为上下两段；通过曲线拟合，得到轮廓线的近似函数表达式，计算该函数一阶导数为零的点，将该点作为人脸侧面特征点。

为了从特定人脸上提取特征，可在人脸模型上定义 41 个特征点。这里特征点可以从特定人的正面和侧面像中提取出来，特征点的自动提取属于人脸面部图象检测与分析范畴，假定已经应用分析与识别技术从特定人脸图象上提取了所需特征或变形曲线，然后将其当作对一般人脸模型的变形参数。由于一般人脸中性模型是一个三维网格体，每个特征点的三维坐标是已知的，在从一般人脸中性模型到特定人脸中性模型的修改过程中，要进行两种变换。首先要对一般人脸中性模型进行整体变换，整体变换的目的是完成面部整体轮廓的修改，使其与特定人的脸形和五官的大致位置相匹配。设变换前的人脸模型上点坐标为 (x, y, z) ，变换后为 (x', y', z') ，变换前后脸的中心点坐标分别为 $o(x_0, y_0, z_0)$ 和 $o'(x'_0, y'_0, z'_0)$ 。其中，脸的中心点 o 定义为两眼眼角点连线及人脸纵向中轴线之间的交点。参数 $p, q1, q2$ 分别定义为中心点到鬓角，中心点到额头中点，中心点到下巴之间的距离；参数 u 定义为嘴部中心点到耳朵下缘距离；参数 $r1, r2, r3$ 在侧面分别定义为额头最高处到发际的距离，中心点到耳朵上缘的距离，嘴角点到耳朵下边缘的距离。

对于上半部分脸（眼睛水平线以上的部分）采用如下修改公式：

$$x' = x_0' + \frac{p'}{p}(x - x_0)$$

$$\begin{aligned}y' &= y_0' + q_1' / q_1 (y - y_0) \\z' &= z_0' + r_1' / r_1 (z - z_0)\end{aligned}$$

对于脸的中间部分和下部可以类似地修改。

5 局部变换中的眼睛区域上点的修改公式如下:

设: (x, y, z) 为变换前眼睛区域点的坐标, (x', y', z') 为变换后眼睛区域点的坐标, 则有

$$x' = ax + by + cz$$

$$y' = dx + ey + fz$$

10 其中, 变量 a, b, c, d, e, f 可通过带入变换前后的三组特征点, 解六元线性方程组得到, 经过这个变换, 可以达到对眼睛微小位置的移动及其形状的改变。

上式看出, 修改仅发生在 x 和 y 方向上, 而没有进行深度方面的修改, 因此不能很好地反映特征人脸的侧面信息。同理可进行眉毛和嘴部的修改。

对鼻子部分的变换公式如下: 设变换前后鼻子区域点的坐标分别为 (x, y, z) 和 (x', y', z') , 鼻子的中心点坐标分别为 (x_0, y_0, z_0) 和 (x'_0, y'_0, z'_0) ,

$$x' = x_0' + p' / p (x - x_0)$$

$$y' = y_0' + q' / q (y - y_0)$$

$$z' = z_0' + r' / r (z - z_0)$$

20 完成了整体变换和局部变换之后, 就得到了一张基本具有特定人脸特征的中性三维人脸网格体。

人脸图像中的唇部模型采用两条抛物线拟合上唇线, 一条抛物线拟合下唇线。选择两个嘴角点, 两上唇抛物线的最高点, 下唇抛物线的最低点, 两上唇抛物线的交点。另外在下唇抛物线上增加两点, 上唇抛物线上增加两点, 在两个嘴角点的连线上增加三组每两个可重合的点。对于张嘴形式, 各对重合点分
25 离, 分别成为上下内唇抛物线上的各点。唇部的外轮廓的抛物线方程满足:

$$y = a(x - b)^2 + c$$

其中,系数 a, b, c 可以通过将已知点的坐标带入上述方程即可解得。

动态的唇动模型利用上述张口模型,描述为五条相互关联起来的抛物线,其中包括上唇两条,下唇一条,上下内唇各一条。在进行参数驱动的唇动合成中,给出纵、横方向上开口的距离,或者给出嘴角点以及上下唇的最高点,即可确定相应的唇部开口状态。

在汉语中,语言中每一个能很自然分辨的语音单位为一个音节,通常一个汉字就是一个音节,一般一个音节是由声母和韵母组成的,发音时声母持续时间很短,然后迅速滑到韵母口型,在汉语拼音中,声母有 19 个,韵母 39 个,韵母又分为单韵母、复韵母和鼻韵母。单韵母发音时,舌位唇形在整个发音过程中不变,所以可以看作一个口型。在汉语发音时常见口型基础上定义几个基本口型,交互地改变口型参数,调整皮肤网格体上嘴部区域上的网格点位置,形成代表基本口型的网格体并预先存储下来。根据上面描述的基本口型,可以衍生出一个韵母口型库,衍生规则如下:

- (1) 单韵母发音已有基本口型与之对应
- (2) 对于复韵母和鼻韵母发音口型,可以拆成多个基本口型的线性组合。

对所有复韵母和鼻韵母,都可以得到对应的口型参数,这样就构成了韵母口型库。合成时,根据汉语拼音,找出声母和韵母对应的口型,然后合成出来,必要时,可以在口型之间进行插值,平滑嘴唇的变化。

最后所应说明的是:以上实施例仅用以说明而非限制本发明的技术方案,尽管参照上述实施例对本发明进行了详细说明,本领域的普通技术人员应当理解:依然可以对本发明进行修改或者等同替换,而不脱离本发明的精神和范围的任何修改或局部替换,其均应涵盖在本发明的权利要求范围当中。

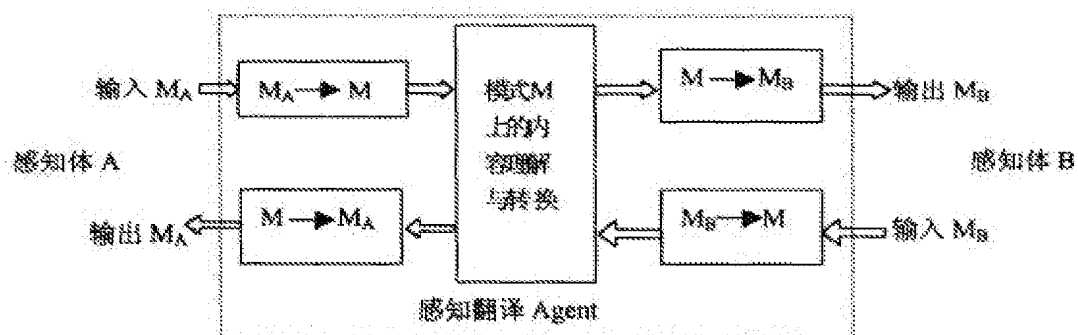


图 1

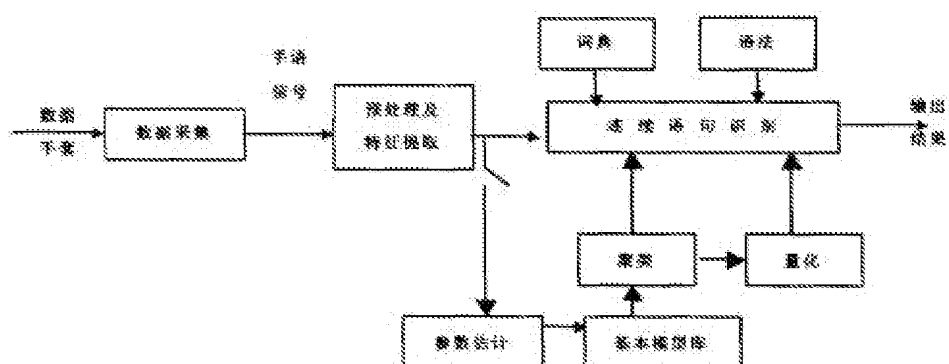


图 2

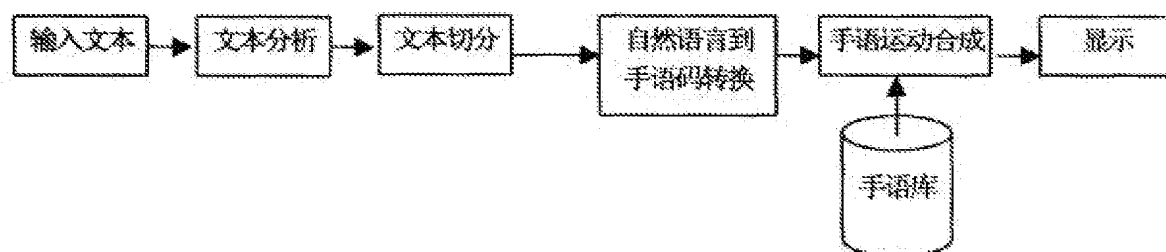


图 3

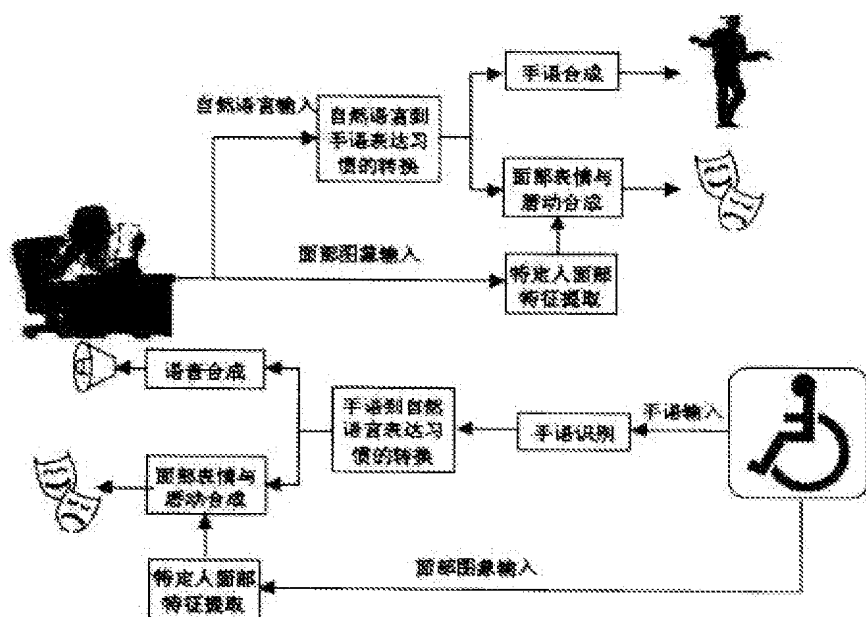


图 4